

Cornfield's Use of Causal Grammar

Milo Schield

Abstract: Cornfield et al. argued that smoking was a causal agent – it had a causal role – in the increase in lung cancer. This argument is arguably the most important usage of an observational association as evidence for causation in the history of statistics. Not only did these authors provide the necessary conditions for a confounder to nullify or reverse an observed association, not only did they review all the counter-claims, not only did they review the supportive evidence (including the dose –response evidence in animals), but they set a standard for the use of grammar in describing the causal status of smoking. This paper reviews their usage of causal grammar.

Background:

Doll and Hill (1952) investigated the etiology of lung cancer. Arkin (1955) noted the association between human smoking habits and death rates. Doll and Hill (1956) provided data on the association between smoking and lung cancer based on reports by UK medical doctors. Fisher (1957) questioned whether this observational association was causal. He provided data showing an association between the degree of twinship (genetics) and smoking preference.

Cornfield et al (1959) reviewed the data available, evaluated the claims of those arguing that this association was not causation and provided a rebuttal to Fisher's argument. In the appendix, they showed that the association between confounder and outcome and between predictor and confounder had to be bigger than that between predictor and outcome in order to nullify or reverse the association. These conditions are the Cornfield conditions.

Schild (1999) analyzed the Cornfield equations and reproduced Cornfield's derivation. Gastwirth, Kreiger and Rosenbaum (2000) reviewed Cornfield's inequality.

In 1964, the Advisory Committee to the Surgeon General of the Public Health Service concluded that "Cigarette smoking is a health hazard of sufficient importance in the United States to warrant appropriate remedial action." The report noted that "Cigarette smoking is *casually related* to lung cancer in men; the magnitude of the effect of cigarette smoking far outweighs all other factors. The data for women, though less extensive, point in the same direction.

Analysis:

The appendix lists all 67 lines using 'caus'. This data is summarized in the following tables.

Table 1: Prevalence of 'Caus' Keywords by Part of Speech (POS)

#	Keyword	#	Keyword	#	Keyword
23	causal (adj)	23	causal (adj)	28	Adjective
18	causes (n)	3	causative (adj)	30	Noun
11	cause (n)	2	noncausal (adj)	8	Verb
7	causes (v)	1	causation (n)	1	Adverb
3	causative (adj)	18	causes (n)		
2	noncausal (adj)	11	cause (n)		
1	causally (adv)	7	causes (v)		
1	causation (n)	1	cause (v)		
1	cause (v)	1	causally (adv)		
67	Total	67	Total	67	Total
Sort keywords by #		Sort keywords by POS		Sort by POS	

Cornfield's Use of Causal Grammar

Milo Schield

Table 2: Prevalence of Keyword Phrases where Word1 is 'Causal (adj)'

Select: Word1 is 'causal' (adj)

#	Word1	Word2	%	Cum%
7	causal	role	32%	32%
6	causal	agent	27%	59%
5	causal	hypothesis	23%	82%
2	causal	relationship	9%	91%
1	causal	effect	5%	95%
1	causal	one	5%	100%
22	Total		100%	

'Because' is excluded since it identifies a reason – a mental cause – for doing something. Their paper focused on physical causes.

Most (59%) of the uses of 'causal' as an adjective involved either 'causal role' or 'causal agent'.

Cornfield et al never asserted that smoking caused cancer without some sort of qualification. But they did argue that smoking was a 'causal agent', that smoking had a 'causal role'. What did these phrases mean?

Here are four possible reasons. They meant:

1. for these phrase to designate an association that strongly supported causation.
2. to use these equivocal phrases in order to avoid taking a stand on the causal status of smoking.
3. for these phrases to indicate that smoking was a probabilistic cause – not a sufficient cause – but still a cause?
4. for these phases to indicate that smoking was one of several factors that might cause lung cancer?

Conclusion:

Cornfield et al., seemed quite willing to take a stand on the causal role of smoking so #2 seems unlikely. If they were upholding #1 then they should have no quarrel with Fisher. But they did have a quarrel with Fisher. Thus it seems that the last two explanations are most plausible.

It seems that any phrase involving some form of 'caus' (excluding 'holocaust' and 'because') should be interpreted as asserting causation.

Bibliography

Cornfield J, Haenszel W, Hammond C, Lilienfeld A, Shimkin M and Wynder E (1959). Smoking and lung cancer: recent evidence and a discussion of some questions. *International Journal of Epidemiology*. 2009;38:1175–1191. www.statlit.org/CP/Cornfield/2009-Int-J-Epidemiol-Reprint-1959-Cornfield.pdf

Doll R, Hill AB (1952). A study of the aetiology of carcinoma of the lung. *Brit M J* 1952;2: 1271–1286.

Gastwirth, J., A. Kreiger and P. Rosenbaum (2000). Cornfield's Inequality. *The Encyclopedia of Epidemiological Methods*. <http://www.statlit.org/CP/Cornfield/2000-Ency-Epi-Methods-Cornfields-Inequality.pdf>

Hill, A. Bradford (1953). Observation and experiment. *New England Journal of Medicine* 248:995-1001.

Schild, M. (1999). Simpson's Paradox and Cornfield's Conditions. *American Statistical Association Proceedings of the Section on Statistical Education*. Copy at www.StatLit.org/pdf/1999SchildASA.pdf

Cornfield's Use of Grammar
Appendix 1. All Phrases Involving 'caus' – Excluding 'because'.

Table 3: Lines involving some form of 'caus'

Page	ID	ID2	Type	Word1	Word2	Context
1175	1	L		causal (adj)	role	tobacco smoking has a causal role in the increase in ...
1175	2	L		causal (adj)	relationship	supports the conclusion of a causal relationship with cigarette
1175	3	R	negative	single	cause (n)	no single cause accounts for all lung cancer
1175	4	L	Quote	causative	factor	that cigarette smoking is a causative factor
1175	5	L		causative	relationship	the inference of a causal relationship between smoking and lung
1176	6	L		causal (adj)	role	deny a causal role of cigarette smoking in cancer of the lung
1177	7	R		sole	cause (n)	that cigarette smoking is the sole cause of lung cancer
1178	8	L		causes (n)	of death	which prevails for many other causes of death
1178	9	L		causation	of	that contribute to the causation of lung cancer
1178	10	R		multiple	causes (n)	convincing evidence for multiple causes of lung cancer.
1178	11	R		sole	cause (n)	untenable to regard smoking of tobacco as the sole cause of lung
1179	12	R		from all	causes (n)	[exclude employees] for all causes who are seriously disabled
1179	13	R		from any	cause (n)	who are seriously disabled from any cause
1179	14	R		other	causes (n)	develop permanent disabilities from disease or other
1179	15	R		from all	causes (n)	account for the deficit in deaths from all causes
1179	16	L		causes (n)	of death	observed for several years for data on mortality and causes of death
1180	17	R		other	causes (n)	If smoking is unrelated to death from lung cancer (or other causes),
1181	18	L		cause (n)	of death	evidence of the cause of death
1182	19	R		multiple	causes (n)	interplay between multiple causes in human pulmonary cancer
1182	20	L		causative	role	the causative role of cigarette smoking
1182	21	L		causal (adj)	role	compatible with the causal role of cigarette smoking in lung cancer.
1183	22	L		causal (adj)	relationship	evidence for the causal relationship of cigarette smoking and lung
1184	23	R		smoking	causes (v)	that cigarette smoking "causes" lung cancer,
1184	24	R		cancer	causes (v)	2) that lung cancer "causes" cigarette smoking,
1184	25	R		common	cause (n)	3) that cigarette smoking and lung cancer both have a common
1185	26	L		causal (adj)	hypothesis	first hypothesis may be referred to as the causal hypothesis
1185	27	L		causal (adj)	hypothesis	entirely consistent with the causal hypothesis
1185	28	L	negative	causally	related	but not causally (adv) related, reduction in the lung-cancer risk.
1186	29	R		common	cause (n)	create a presumption that it is a common cause.
1186	30	R		postulated	cause (n)	proposed as a measure of the postulated cause
1186	31	L	negative	noncausal	nature	1) appraising the possible noncausal nature of an agent
1186	32	R		to be	causal (adj)	appraising the public health significance of an effect known to be
1186	33	L	negative	causal (adj)	effect	If an agent, A, with no causal effect upon the risk of a disease,
1186	34	L		causal (adj)	agent	because of a positive correlation with some other causal agent
1186	35	L		causal (adj)	agent	not because cigarette smoke in [is] a causal agent
1187	36	R		real	causes (n)	The presence of other real causes thus reduces the apparent relative
1187	37	R		other	causes (n)	the same in the absence of other causes
1187	38	L		causal (adj)	agent	If a causal agent A increases the risk for disease I
1187	39	L		causal (adj)	hypothesis	The causal Hypothesis
1187	40	L		causal (adj)	hypothesis	consideration of some of the contradictions in the causal hypothesis,
1187	41	R	quote	the	cause (n)	that increasing smoking is 'the cause'
1187	42	L		causal (adj)	hypothesis	Far from making the causal hypothesis untenable
1187	43	L		causes (n)	of death	but from many other causes of death
1187	44	L	quote	causes (n)	of all ills	that cigarette smoking is one of the causes of all ills
1187	45	R	quote	such	causes (n)	this rate includes such causes as accident, homicide, etc.
1187	46	R	negative	not a	cause (n)	cigarette smoking is a symptom, not a cause
1187	47	R	other	accidental	causes (n)	only with respect to non-accidental causes of death
1187	48	L	other	causal (adj)	one	some other explanation than a causal one
1187	49	R		should	cause (v)	find it quite incredible that smoking should cause all these diseases
1187	50	R	indent	smoking	causes (v)	that smoking causes lung cancer
1187	51	R	indent	smoking	causes (v)	that smoking causes or provokes a whole gamut of diseases
1188	52	R		number of	causes (n)	increased the death rate for a number of causes
1188	53	L		causal (adj)	role	a reason for doubting the causal role of the Fog.
1188	54	L		cause (n)	and effect	universe in which cause and effect
1188	55	L		causal (adj)	agent	1) The evidence that tobacco is a causal agent
1188	56	L		causal (adj)	role	we could not exclude the possibility that cigarettes play a causal role
1188	57	L		causal (adj)	role	2) Accepting, for the sake of discussion, the causal role of cigarettes
1188	58	R		other	causes (n)	presence of other causes will be manifested
1188	59	R	hypothetical	consumption	causes (v)	even if cigarette consumption causes an elevation
1188	60	R		other	causes (n)	other causes of great importance must also be present
1188	61	R		other	causes (n)	that there are other causes of lung cancer
1188	62	R	accepted	smoke	causes (v)	that tobacco smoke causes thromboangiitis obliterans
1188	63	R	accepted	it	causes (v)	it causes tumors when painted on the backs of susceptible mice
1188	64	L		causal (adj)	agent	fail to develop lung cancer if cigarettes are in fact a causal agent?
1188	65	L		causal (adj)	role	not a reason, however, for doubting the causal role
1191	66	L		causal (adj)	agent	Let the disease rate for those exposed to the causal agent, B, be r1
1191	67	L	negative	noncausal	agent	unaffected by exposure or nonexposure to the noncausal agent, A.

Cornfield's Use of Grammar
Appendix 1. All Phrases Involving 'caus' – Excluding 'because'.

Table 4: Sort 'caus' phrases by ID2, Keyword, Word1, Word2

ID	ID2	Keyword	Word1	Word2	ID	ID2	Keyword	Word1	Word2
34	L	causal (adj)	causal (adj)	agent	32	R	causal (adj)	to be	causal (adj)
35	L	causal (adj)	causal (adj)	agent	25	R	cause (n)	common	cause (n)
38	L	causal (adj)	causal (adj)	agent	29	R	cause (n)	common	cause (n)
55	L	causal (adj)	causal (adj)	agent	13	R	cause (n)	from any	cause (n)
64	L	causal (adj)	causal (adj)	agent	46	R	cause (n)	not a	cause (n)
66	L	causal (adj)	causal (adj)	agent	30	R	cause (n)	postulated	cause (n)
33	L	causal (adj)	causal (adj)	effect	3	R	cause (n)	single	cause (n)
26	L	causal (adj)	causal (adj)	hypothesis	7	R	cause (n)	sole	cause (n)
27	L	causal (adj)	causal (adj)	hypothesis	11	R	cause (n)	sole	cause (n)
39	L	causal (adj)	causal (adj)	hypothesis	41	R	cause (n)	'the	cause' (n)
40	L	causal (adj)	causal (adj)	hypothesis	49	R	cause (v)	should	cause (v)
42	L	causal (adj)	causal (adj)	hypothesis	47	R	causes (n)	accidental	causes (n)
48	L	causal (adj)	causal (adj)	one	12	R	causes (n)	from all	causes (n)
22	L	causal (adj)	causal (adj)	relationship	15	R	causes (n)	from all	causes (n)
21	L	causal (adj)	causal (adj)	role	10	R	causes (n)	multiple	causes (n)
53	L	causal (adj)	causal (adj)	role	19	R	causes (n)	multiple	causes (n)
56	L	causal (adj)	causal (adj)	role	52	R	causes (n)	number of	causes (n)
57	L	causal (adj)	causal (adj)	role	14	R	causes (n)	other	causes (n)
65	L	causal (adj)	causal (adj)	role	17	R	causes (n)	other	causes (n)
2	L	causal (adj)	causal (adj)	relationship	37	R	causes (n)	other	causes (n)
1	L	causal (adj)	causal (adj)	role	58	R	causes (n)	other	causes (n)
6	L	causal (adj)	causal (adj)	role	60	R	causes (n)	other	causes (n)
28	L	causally (adv)	causally (adv)	related	61	R	causes (n)	other	causes (n)
9	L	causation (n)	causation (n)	of	36	R	causes (n)	real	causes (n)
4	L	causative (adj)	causative (adj)	factor	45	R	causes (n)	such	causes (n)
5	L	causative (adj)	causative (adj)	relationship	24	R	causes (v)	cancer	causes (v)
20	L	causative (adj)	causative (adj)	role	59	R	causes (v)	consumption	causes (v)
54	L	cause (n)	cause (n)	and effect	63	R	causes (v)	it	causes (v)
18	L	cause (n)	cause (n)	of death	62	R	causes (v)	smoke	causes (v)
44	L	causes (n)	causes (n)	of all ills	23	R	causes (v)	smoking	causes (v)
8	L	causes (n)	causes (n)	of death	50	R	causes (v)	smoking	causes (v)
16	L	causes (n)	causes (n)	of death	51	R	causes (v)	smoking	causes (v)
43	L	causes (n)	causes (n)	of death					
67	L	noncausal (adj)	noncausal (adj)	agent					
31	L	noncausal (adj)	noncausal (adj)	nature					